

The evolutionary history of chromosomal super-integrans provides an ancestry for multiresistant integrans

Dean A. Rowe-Magnus*, Anne-Marie Guerout*, Pascaline Ploncard*, Broderick Dychinco†, Julian Davies†, and Didier Mazel**

*Unité de Programmation Moléculaire et Toxicologie Génétique, Centre National de la Recherche Scientifique Unité de Recherche Associée 1444, Département des Biotechnologies, Institut Pasteur, 25 Rue du Dr. Roux, 75724 Paris, France; and †Department of Microbiology and Immunology, University of British Columbia, 6174 University Boulevard, Vancouver, BC, Canada V6T 1Z3

Communicated by Allan Campbell, Stanford University, Stanford, CA, November 2, 2000 (received for review August 1, 2000)

Integrans are genetic elements that acquire and exchange exogenous DNA, known as gene cassettes, by a site-specific recombination mechanism. Characterized gene cassettes consist of a target recombination sequence (*attC* site) usually associated with a single open reading frame coding for an antibiotic resistance determinant. The affiliation of multiresistant integrans (MRIs), which contain various combinations of antibiotic resistance gene cassettes, with transferable elements underlies the rapid evolution of multidrug resistance among diverse Gram-negative bacteria. Yet the origin of MRIs remains unknown. Recently, a chromosomal super-integron (SI) harboring hundreds of cassettes was identified in the *Vibrio cholerae* genome. Here, we demonstrate that the activity of its associated integrase is identical to that of the MRI integrase, *IntI1*. We have also identified equivalent integron superstructures in nine distinct genera throughout the γ -proteobacterial radiation. Phylogenetic analysis revealed that the evolutionary history of the system paralleled that of the radiation, indicating that integrans are ancient structures. The *attC* sites of the 63 antibiotic-resistance gene cassettes identified thus far in MRIs are highly variable. Strikingly, one-fifth of these were virtually identical to the highly related yet species-specific *attC* sites of the SIs described here. Furthermore, antimicrobial resistance homologues were identified among the thousands of genes entrapped by these SIs. Because the gene cassettes of SIs are substrates for MRIs, these data identify SIs as the source of contemporary MRIs and their cassettes. However, our demonstration of the metabolic functions, beyond antibiotic resistance and virulence, of three distinct SI gene cassettes indicates that integrans function as a general gene-capture system for bacterial innovation.

The impact of lateral gene transfer on bacterial evolution is underscored by the realization that foreign DNA can represent up to one-fifth of a given bacterial genome (1). Perhaps the most striking embodiment of its affect on microbial adaptation has been the rapid and widespread emergence of similar antibiotic-resistance profiles among phylogenetically diverse Gram-negative clinical and environmental isolates over the last half-century (2). The localization of antibiotic-resistance determinants to mobile entities such as plasmids and transposons readily explained this phenomenon (3–6). Closer examination revealed that in many cases a new type of genetic element, termed an integron, harbored the resistance determinants. Integrans are natural cloning and expression systems that incorporate open reading frames and convert them to functional genes (for review see refs. 7 and 8). The integron platform codes for an integrase (*intI*) that mediates recombination between a proximal primary recombination site (*attI*) and a secondary target called an *attC* site [or 59-base element (59be)]. The *attC* site is normally found associated with a single open reading frame (ORF), and the *attC*-ORF structure is termed a gene cassette (9–12). Insertion of the cassette at the *attI* site, which is

located downstream of a resident promoter internal to the *intI* gene, drives expression of the encoded proteins (13). Most of the *attC* sites of the integron cassettes identified to date are unique. Their length and sequence vary considerably (from 57 to 141 bp) and their similarities are primarily restricted to their boundaries, which correspond to the inverse core site (RYYAAC) and the core site [G↓TTRRRY; ↓, recombination point (11, 14)]. More than 60 different antibiotic-resistance genes, covering most antimicrobials presently in use, have been characterized in cassette structures (15). The stockpiling of exogenous genetic loci to create multiresistant integrans (MRIs) has contributed substantially to the current dilemma in the treatment of infectious disease (2, 8, 15, 16), as MRIs harboring up to five different resistance cassettes have been described (17). However, the origin of integrans and their cassettes has remained obscure.

Recently a distinct type of integron was discovered in the *Vibrio cholerae* genome (18). Located on the smaller of the two circular chromosomes (19), this element spans 126 kb and gathers at least 179 cassettes (20, 21) into a single structure termed a super-integron (SI), dwarfing previously identified MRIs. Of particular interest were the observations that: (i) the cassette-associated *attC* sites, termed VCRs [for *Vibrio cholerae* repeats (22); other repeated sequences are named similarly] displayed a strikingly high degree of sequence relatedness, unlike their counterparts of MRIs; (ii) the *attC* site of two MRI antibiotic-resistance gene cassettes, *CARB4* and *dfvI*, was similar to the VCRs; and (iii) the *V. cholerae* SI cassettes were substrates for the class 1 integrase of MRIs (18).

These relationships led us to hypothesize that MRIs evolved from SIs through the entrapment of *intI* genes and their cognate *attI* sites into highly mobile structures like transposons. Subsequent harvesting of cassettes from various SI sources then led to the establishment of contemporary MRIs. In this scenario, each distinct *attC* site of MRI cassettes would represent a specific SI cluster containing related *attC* sites, as in *V. cholerae*. If SIs are the progenitors of MRIs and their cassettes, then the SI integrases should be active, the cassettes should encode adaptive functions, and SIs should be numerous in the bacterial kingdom. Here, we have demonstrated that the *V. cholerae* SI integrase is active for site-specific cassette recombination, and we have

Abbreviations: MRI, multiresistant integron; SI, super-integron; VCR, *Vibrio cholerae* repeated sequence (repeated sequences of other species are named similarly); IS, insertion sequence; Tn, transposon.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF180939, AY014398–AY014401, AF324210, AF324211, AF324482–AF324484, and AF324934–AF324946).

*To whom reprint requests should be addressed. E-mail: mazel@pasteur.fr.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. VchIntIA-dependent recombination frequencies for the CARB4 and ORF1::cat cassette substrates

Strain	Plasmid	Recombination frequency
ω 32	pSU38::CARB4, pVC3	5.0×10^{-3}
ω 34	pSU38::CARB4	$<10^{-6}$
ω 33	pSU38::ORF1-cat, pVC3	1.6×10^{-2}
ω 35	pSU38::ORF1-cat	$<10^{-5}$

determined the metabolic function of three SI cassettes: a sulfate-binding protein, a psychrophilic lipase, and a restriction enzyme. Furthermore, a systematic search for SI structures beginning with *Vibrio* and closely related species and extending to distantly related genera has revealed that this gene acquisition machinery is an ancient system that is widespread among the proteobacteria.

Materials and Methods

Bacterial Strains, PCR, Sequencing, and Phylogenetic Analysis. Bacterial strains were provided by the Collection de l'Institut Pasteur (CIP). All genomic DNA was isolated by using the Qiagen DNEasy Kit. PCR-amplified genes were cloned by using the TA TOPO cloning kit (Invitrogen) and verified by sequencing of the corresponding genomic clones [performed by MWG-Biotech (Ebersberg, Germany), on a Pharmacia 4 × 4 sequencing system or an Applied Biosystems 377 sequencer]. Primers were obtained from Genset (Evry, France) or MWG-Biotech. Alignments were generated with the CLUSTALX (23) software program (version 1.8) and dendrograms were compiled by using the neighbor-joining method (computed from 1,000 independent trials) of CLUSTALW and PHYLIP (24) or TREEVIEW (25). 16S rRNA sequences were retrieved from GenBank.

Integrase Activity and Cassette Function Assays. The VchIntIA gene (Vch indicating the *V. cholerae* source; other genes are identified similarly) was digested with *Eco*RI and *Bam*HI after PCR amplification with the primers VINTup (CCGGAATTCTGTACAAATAACCAAGTTAAAT) and VINTdw (CGCGGATCCTGTATTATTAAGAAAGGGGAG) and cloned into pUC18 digested with the same enzymes to create pVC3. The substrate plasmids pSU38::CARB4 and pSU38::ORF1-cat have been previously described (18). *Escherichia coli* DH5 α (*recA*[−]) containing the indicated plasmids was incubated overnight at 37°C with the appropriate antibiotics. Plasmid DNA was prepared and digested with either *Xba*I (for pSU38::ORF1-cat) or *Mfe*I (for pSU38::CARB4), which cut specifically within the respective cassettes. Samples of both the original and digested plasmid preparations were used to transform *E. coli* DH5 α . The recombination frequency is the number of ampicillin-sensitive (ω 32 and ω 34) or chloramphenicol-sensitive (ω 33 and ω 35) transformants among the number of kanamycin-resistant transformants (average of two independent trials; Table 1). The precise location of the recombination event was established by PCR and sequencing of 25 random clones per test set to determine the VchIntIA-dependent recombination frequency.

The sulfate-binding protein cassette of the *V. cholerae* SI was amplified by PCR with the primers VCR1/SBP (CGTGAATCTTTGAGGCGTTTGTAGGTTT) and SBP3' (ATAGGATCCTCATTTTTAACTACCTTTTAC), digested with *Eco*RI and *Bam*HI, and cloned into pUC18 digested with the same enzymes to create pVchSBP. Complementation of the sulfate auxotroph *E. coli* *shp/cysP* double mutant, EC2402 (26), was carried out in liquid and solid minimal media containing 0.1 mM sodium sulfate as the sole source of sulfate. The lipase cassette of the *Vibrio marinus* SI was amplified by PCR with the primers Vmarlip1 (TCTAGACAAACGTGTTATGCAG-

GCGTT) and Vmarlip3 (GAATTCCTTATGGTTTAAATGC-GAAAGTTAAAAAG). The product was digested with *Nde*I and *Bam*HI and cloned into pET3a digested with the same enzymes to give pET3aVml. T7 RNA polymerase-dependent lipase expression in recombinant *E. coli* BL21(DE3)pLysS was induced by the addition of isopropyl β -D-thiogalactoside (IPTG) to a final concentration of 0.5 mM on solid LB medium containing 100 μ g/ml ampicillin, 25 μ g/ml chloramphenicol, and 0.5% tributyrin as previously described (27). Positive clones produced a zone of clearing caused by hydrolysis of the lipid in the surrounding medium. Control strains carried the vector without insert.

Identification of the Integrase Genes. A primer to *rpmI* (L35), R16–1 (CCAAACGCCATTCTGCCTAAG), and VCR1 (18) were used to amplify VmintIA from *Vibrio mimicus* genomic DNA. Primers VCR1 and H7–1 (GGACTATGGCAAGTTCCTCT), which targeted *infC* (IF3), were used to amplify the entire VmeintIA gene from *Vibrio metschnikovii* chromosomal DNA. The VmeintIA gene was used to probe a *Listonella pelagia* HindIII library in pNot218R [modified from the phagemid pTZ18R vector (Pharmacia) to include in-frame *Not*I sites on both sides of the multiple cloning site]. Single-strand DNA was prepared after infection with M13K07 helper phage. This library was used in a PCR with the M13 forward primer and a set of degenerate *intl* primers, 17–2C [C(AC)A(AG)(CAT)A-C(AG)TG(CTA)GTGTAGAT]. The PCR product, corresponding to LpeintIA, was cloned and sequenced. This gene was in turn used as a probe to clone a 6-kb *Vibrio parahaemolyticus* fragment containing VpaintIA from an *Xba*I library in pNot218R.

On the basis of estimates of the size of the *Vibrio* SIs, it was assumed that the *Vibrio fischeri* SI would be more or less equal in length (ca. 3% of the genome). Fifty random clones of a *V. fischeri* BglII genomic DNA library (in pNot218R) were selected for sequencing, and one was found to contain a short cassette array plus chromosomal DNA with homology to general house-keeping genes, indicative of a clone corresponding to the 3' end of the SI. Primers VFR1 (GCAATCGTGTTGATGCATTTGTTA) and VFR2 (CAAACGTGTTATGCAGGCGTTA), designed from this sequence, were used in a PCR screen to determine the consensus sequence of the VFRs from several independent clones containing nonoverlapping cassette arrays. A unique conserved restriction enzyme site, *Nsi*I, was identified among the VFRs near the 5' end. *V. fischeri* genomic DNA was digested with *Nsi*I, with the expectation that the enzyme would cut regularly between most of the SI cassettes but less frequently in the rest of the genome. The majority of DNA fragments that hybridized with a VFR-specific probe were less than 2.5 kb, but three larger hybridizing fragments of 3, 4.5, and 6 kb, containing at least one VFR were observed. Rehybridization with the 3' end of the SI as a probe gave an intense signal corresponding to the 3-kb fragment. We surmised that the remaining two fragments likely corresponded to either the 5' end of the SI or cassette arrays in which the enzyme did not cut. *Nsi*I partial libraries corresponding to the 4.5- and 6-kb regions were constructed in *Nsi*I-digested pMTL22 (28). Because the *Nsi*I site was located close to the 5' end of the VFR, unique fusion primers, NsiFUS1 (GTACCATGGCATGCATTTGTT) and NsiFUS2 (GAGATCTCTCGATGCATTTGTT), corresponding to a region of the multiple cloning site of pMTL22 immediately adjacent to and including the *Nsi*I site plus an additional 5 nt specific to the 5' end of the VFR sequence were created. These were used in a PCR with VFR2 and the respective libraries as matrix to identify the proper clones. Subcloning and sequencing of the 6-kb fragment unveiled VfiintIA.

Analysis of the available sequencing data [The Institute for Genomic Research (TIGR)] revealed an integron-integrase gene in the genome of *Shewanella oneidensis* MR-1 (formerly

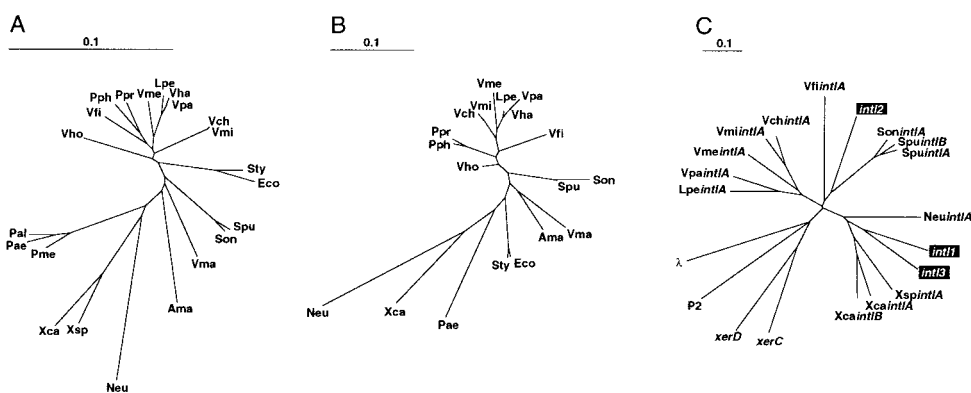


Fig. 1. Phylogenetic relationship of the proteobacteria. Unrooted dendrograms based on the 16S rRNA (A), *rplT* (ribosomal L20) (B), and *intI* (C) genes among the proteobacteria used in this study. The integrases from the three classes of MRIs are boxed in C. Organism abbreviations are as follows: Vch, *Vibrio cholerae*; Vme, *V. metschnikovii*; Vpa, *V. parahaemolyticus*; Vma, *V. marinus*; Vha, *V. harveyi*; Vho, *V. hollisae*; Vfi, *V. fischeri*; Lpe, *Listonella pelagia*; Pph, *Photobacterium phosphoreum*; Ppr, *P. profundum*; Eco, *Escherichia coli*; Ave, *Aeromonas veronii*; Son, *Shewanella oneidensis*; Spu, *S. putrefaciens*; Xca, *Xanthomonas campestris*; Xsp, *Xanthomonas sp.*; Neu, *Nitrosomonas europaea*. For the integrase and recombinase proteins: λ (of phage lambda), P2 (of phage P2), XerC/D (of *E. coli*). For *intI* nomenclature see Table 2.

Shewanella putrefaciens MR1). However, mobile elements [insertion sequences (ISs) or transposons (Tns)] had disrupted the SI structure. A native SI structure in strain 69.34 was identified as follows: Primers SpintI-1 (GACAACCAGCAAGGGAT-AGGT) and SpintI-3 (TTTATAATGTGTCTAACGGGC), based on the integrase gene of *S. oneidensis* MR1 (SonintI_A), were used to amplify the 3' end of SpuintI_A of *S. putrefaciens* strain 80.40T. Again, the SI of this strain was also disrupted by IS or Tn elements. Using SpuintI_A as a probe, we identified SpuintI_B of the *S. putrefaciens* strain 69.34 SI by Southern analysis and sequenced it from a *Bam*HI library in pNot218R.

A search of the GenBank database with the *attC* site of the *aadA7* aminoglycoside-resistance gene cassette (29) revealed that the *Xba*I restriction-modification locus of *Xanthomonas campestris* pv. *badrii* was coded in a cassette-like structure. SIs were identified in the genomes of several *X. campestris* pathovars and *Xanthomonas* spp. as follows: The primers XCR1 (CGTCG-GCTTGAATGAATTGTTA) and XCR2 (GCGCGGCTTA-ACTCAGGTGTT), designed from the *attC* site of *aadA7* were used to PCR amplify several cassettes in *X. campestris* pv. *campestris* 100069T. These cassettes were used as probes to clone a 26-kb *Eco*RI genomic fragment, containing a large portion of the SI cassette array less the integrase gene, in pUC18. The *XcaintI_A* gene was obtained by sequencing of an overlapping 12-kb *Pst*I fragment. The *intI* gene of *Xanthomonas* sp. 102397 (XspintI_A) and *X. campestris* pv. *badrii* (XcaintI_B) were cloned by PCR with the primers XCR1 and intI9RHS (ATGCGTG-GCGAACGAATGCCG) designed from the *XcaintI_A* gene.

SI Signature Detection Among Various Bacterial Genera. SI structures were identified in various species by PCR amplification of VCR-like cassettes with primers designed to detect the XXRs of closely related species as described above (*Vibrio marinus*, *Vibrio anguillarum*, *Vibrio salmonicida*, and *Photobacterium phosphoreum* with VCR1 and VCR2 or VFR1 and VFR2) or by Southern hybridization with *intI* or cassette specific probes (*Vibrio hollisae*, *Vibrio harveyi*, *Alteromonas macleodii*, *Pseudomonas pseudoalcaligenes*). All PCR products were verified as sharing the characteristics of typical integron gene cassettes (8, 21).

Cloning of the *rplT* Genes. The *rplT* genes of the following species were cloned after PCR amplification with the indicated set of primers. Degenerate positions are bracketed: *V. mimicus* and *V. metschnikovii* with primers VM20.SOS (CGTCAACTACGC-

CCTAACTCA) and VM20.TPT (CTTACGACGATCTC-GATAGAT); *S. putrefaciens* strains 80.40T and 69.34 with primers SpL20-1 (ATATGCTTACCGTGACCGTCG) and SpL20-2 [(AT)(AG)CTTTGTCTGAATAC(N)GCGAT]; *V. marinus* and *A. macleodii* with primers SputL20-1 [ATA(CT)GCT-TACCGTGACCGTCG] and SputL20-3 [(AT)(AG)CTTTGT-CAAATAC(N)GCGAT]; *P. phosphoreum*, *Photobacterium profundum*, *L. pelagia*, *V. fischeri*, *V. harveyi*, *V. parahaemolyticus*, and *V. hollisae* with primers DRL20U (GCCTCGCGTA-AAACGTGGTGT) and DRL20D (TTTGTCTGAATACTGC-GATATC); *X. campestris* with primers XcL20-2 [TTTGTCTGAATAC(GT)GCGATATC] and XcL20-4 [(TAG)TG-GAT(CT)GC(TAG)CGTATCAACGC].

Deposition of Nucleic Acid Sequences. All nucleic acid sequences have been deposited at the National Center for Biotechnology Information (NCBI) GenBank Database. Accession numbers can be found on our website: <http://www.pasteur.fr/recherche/unites/pmtg/>.

Results

A Conserved Activity for the Integron-Integrases. To verify the evolutionary link between MRIs and the *V. cholerae* SI, an activity for the SI integrase gene, *intI₄* (renamed *VchintI_A*, see below), had to be demonstrated. We monitored integrase activity through deletion of a selectable marker caused by *VchintI_A*-dependent site-specific recombination. We used as substrates pSU38::ORF1-cat, which harbors a chloramphenicol-resistance gene bordered by two VCRs, and pSU38::CARB4, which contains a carbenicillinase-resistance gene cassette between two *attC* sites of an MRI (18). The junctions of 25 recombinant plasmids from each test set were sequenced to ensure that recombination occurred specifically within the GTT triplet of the core site. The results, presented in Table 1, clearly indicated the presence of a *VchintI_A*-dependent site-specific recombination activity for both cassette substrates that was congruent with the activity of IntI1 (18), the only other integron-integrase activity demonstrated to date. Furthermore, the residues identified as critical for the recombination activity of IntI1 (30) were absolutely conserved in *VchintI_A* (see Fig. 3, which is published as supplemental data on the PNAS web site, www.pnas.org), suggesting that both catalyze recombination by the same mechanism.

Table 2. Attributes of the chromosomal SIs

Strain	Size of SI, kb*	Integrase gene	Length of <i>intI</i> , nt	Name	Length, nt	XXR	
						Similarity	
						% identity	No. cassettes†
<i>V. cholerae</i> 569B	>100	VchintIA	963	VCR	121–123	74	23
<i>V. mimicus</i> 101888T	>100	VmiintIA	963	VMiR	121–123	80	3
<i>V. metschnikovii</i> A267	>100	VmeintIA	963	VMeR	117–120	83 [‡]	20 [‡]
<i>V. parahaemolyticus</i> 75.2T	>100	VpaintIA	963	VPR	119–120	75 [‡]	4 [‡]
<i>V. fischeri</i> 103206T	>100	VfiintIA	951	VFR	119–120	71 [‡]	16 [‡]
<i>L. pelagia</i> 10276.2T	>100	LpeintIA	963	LPR	117–120	86	3
<i>S. oneidensis</i> MR1	ND	SonintIA	960	ND	ND	ND	ND
<i>S. putrefaciens</i> 80.40T	ND	SpuintIA	665 [§]	ND	ND	ND	ND
<i>S. putrefaciens</i> 69.34	ND	SpuintIB	954	SPR	68–94	ND	ND
<i>X. campestris</i> pv. <i>campestris</i> 100069T	ND	XcaintIA	1,014	XCcR	53–54	70	10
<i>X. campestris</i> pv. <i>badrii</i> 11672	ND	XcaintIB	285 [§]	XCbR	53–54	—	—
<i>X. sp.</i> 102397	ND	XspintIA	858 [§]	XSR	53–54	80	7
<i>N. europaea</i> 103999	ND	NeuintIA	975	ND	ND	ND	ND

ND, not determined.

*> signs denote estimations by Southern blot with the respective XXR-specific probes.

†Number of cassettes sequenced.

‡Two distinct subgroups with at least 60% identity between them were present in the total number of cassettes examined. The percent similarity is among XXR within the respective subgroups.

§Partial sequence of *intI* gene.

SI Distribution. Among the Vibrionaceae and their close relatives (Fig. 1), we have identified SI structures in *V. mimicus*, *V. metschnikovii*, *V. parahaemolyticus*, *L. pelagia* and *V. fischeri*. Estimates by DNA hybridization revealed that each of the *Vibrio* species SI were at least 100 kb, collectively equaling a small genome (for example, *Mycoplasma genitalium*). Sequencing of a portion of these SI ($\approx 10\%$) indicated that, as seen for *V. cholerae* (21), the majority of the cassette-borne genes had no counterpart in the database, or the sole homologues were unassigned ORFs of viral, bacterial, or eukaryotic origins. Several distinct genera of the γ -proteobacteria also harbor SI structures, including *Shewanella* and *Xanthomonas*. In addition, SIs were recently characterized in *Pseudomonas alcaligenes* and *Pseudomonas mendocina*.[§] SI signatures were obtained by PCR for a gamut of γ -proteobacteria, including *V. harveyi*, *V. marinus*, *V. hollisae*, *V. anguillarum*, *V. salmonicida*, *Alteromonas macleodii*, *Photobacterium phosphoreum*, *Moritella marina*, several *Xanthomonas* spp. (Collection de l'Institut Pasteur nos. 102336, 102338, and 105155), and *Pseudomonas pseudoalcaligenes*. Notably, integron-integrase genes were also identified by blast analysis in the genome of *Nitrosomonas europaea*, a member of the β -proteobacterial radiation (U.S. Dept. of Energy sequencing program).

For each SI identified (Table 2), an *intI* gene was found to be divergently transcribed from an upstream gene-XXR cluster (in which the first X of XXR denotes the genus and the second X denotes the species in which the repeat occurs). This structure was identical to previously identified MRIs and the *V. cholerae* SI (8). Because of the large number of integrases that were found, a species-letter nomenclature system for the *intI* genes was adopted. For example, the integrase gene of *V. cholerae* 569B was designated VchintIA and the integrase genes of *S. putrefaciens* strains 80.40T and 69.34 were designated SpuintIA and SpuintIB, respectively.

Integrans Are Ancient Structures. Dendrograms were constructed to probe the evolutionary relationship between the extent of

species and *intI* gene divergence. These were based on two essential chromosomal markers unlikely to be subject to horizontal transfer (31), the 16S rRNA gene and the *rplT* gene (coding for the ribosomal L20 protein). Comparison of the 16S rRNA, *rplT*, and *intI* trees revealed a robust phylogenetic congruence for the distribution of the species used in this study (Fig. 1). The extent of divergence between the *intI* genes adhered to the line of descent among the bacterial species. Consequently, the integrases partitioned in genus-specific clades. The *V. cholerae* and *V. mimicus* integrases clustered together, whereas the integrases of *V. metschnikovii*, *V. parahaemolyticus*, *L. pelagia*, and *V. fischeri*, species on separate lines of descent from *V. cholerae* according to their 16S rRNA sequences, were more distantly related. This superimposable dispersal pattern was also evident for the integrase genes of the *Shewanella* and *Xanthomonas* genera. SonintIA, SpuintIA, and SpuintIB formed a clade, whereas the integrase genes of *X. campestris* pv. *campestris*, *X. campestris* pv. *badrii*, and *Xanthomonas* sp. CIP 102397 partitioned in a separate clade. The congruence of the three dendrograms argued for the presence of an integron in the ancestral organism of each genus, and perhaps the γ -proteobacterial radiation itself, over the independent acquisition of integron platforms (i.e., the *intI* gene and the *attI* site) within each lineage. Furthermore, no evidence was found that the SI platforms were mobile. Thus integrases are ancient structures that have been steering the evolution of bacterial genomes for hundreds of millions of years, as the substitution rates calculated by Ochman and Wilson estimate that the *Vibrio* and *Pseudomonas/Xanthomonas* genera separated some 300–800 million years ago (32).

Despite their divergence, the integron-integrase genes clearly formed a related but distinct group from the λ family of site-specific recombinases (Fig. 1C). Furthermore, the *Shewanella* integrase genes clustered with *intI2* and the *Xanthomonas* integrases grouped with *intI1* and *intI3*, suggesting that the root for the three classes of MRIs may be within these divisions. Because the catalytic residues were absolutely conserved among all of the SI integrases identified, it can be presumed that they all use the same mechanics for cassette

[§]Vaisvila, R., Morgan, R. & Raleigh, E., Ninth International Congress of Bacteriology and Applied Microbiology, Aug. 15–20, 1999, Sidney, Australia, abstr. BP09.27.

directional flow of cassettes from SIs to MRIs. If SIs are the source of MRI cassettes, the gene cassette pool available to be propagated through mobile integrons is immense.

Our data suggest several ideas regarding integron-mediated lateral transfer. First, that each distinct resistance cassette *attC* site of MRIs represents a specific SI. Second, that MRIs have evolved from SIs through the association of *intI* genes and their cognate *attI* sites with mobile structures. Subsequently, cassettes with dissimilar *attC* sites were harvested from the many different kinds of SI cassette pools through multiple lateral transfers. Third, as illustrated by the grouping of VCRcA with the VMeRs, the exchange of cassettes between SIs, whether direct or indirect, must also occur. It has been noted that XXRs occur exclusively in intergenic regions. An XXR primer/reverse transcriptase model for cassette genesis has been proposed to accommodate this observation (36), but several characteristics of SIs (i.e., the apparent varied origins of the genes, promoter-associated cassettes, and inversely oriented genes) are inconsistent with this hypothesis. The homogeneous yet distinctive nature of the XXRs is an attribute that may reflect an enzymatic or gene-conversion type event governing their propagation. However, the dynamics of cassette genesis remain somewhat nebulous.

Strikingly, one-fifth of the antimicrobial cassettes of MRIs have *attC* sites virtually identical in length and sequence to the XXRs of the SIs described here. Because the XXRs of SI cassettes are substrates for the class 1 integrase of MRI (18), this implies that these resistance cassettes originated within these or closely related bacteria. Although a known antibiotic-resistance gene cassette has yet to be definitively identified within these SIs, several potential progenitor cassettes with significant homology (greater than 30% amino acid identity) to aminoglycoside, phosphotricin, fosfomycin, streptothricin, and chloramphenicol resistance genes have been found. The function of these cassettes remains to be established. However, the phenomenal chemical diversity of naturally occurring antimicrobial compounds (37) and the effects of perpetual evolutionary changes renders precise determination of their function a daunting task.

The potency of the integron system lies in its versatility, that is, the ability to recognize highly variable target recombination sequences and an apparently limitless capacity to exchange and stockpile cassettes. Such flexibility permits rapid adaptation to the unpredictable flux of environmental niches by allowing bacteria to scavenge foreign genes that may ultimately increase their fitness. In the course of natural selection, the acquisition of *shp* or *lip* genes may promote the colonization of new ecological niches, whereas a restriction-modification system may be a defensive measure against, for example, phage infection. Likewise, genes that fail to provide a meaningful function can be readily eliminated. In addition

to the plethora of antibiotic-resistance genes, two virulence genes of *V. cholerae* are also structured as gene cassettes (38, 39), underscoring the potential of this system to participate in the establishment of pathogenicity islands. Although retained among many extant proteobacterial genera, this system may have been lost in others (for example, the enterobacteria). Genomes are dynamic, and ample opportunities exist for extensive changes in gene order and content to occur. These events can lead to deletion of nonessential genes. Integrons may not be essential systems, and the SI cassettes examined to date seemingly encode adaptive rather than indispensable functions. Thus, deletion of the integron platform may have occurred in certain genera or independently within species. In support of this possibility, one of the *Shewanella* strains we examined was found to contain only a partial *intI* gene. Furthermore, other elements exist that allow bacteria to survive adaptive challenges and explore new niches. The participation of bacteriophages in this regard is well documented (for review see refs. 1, 40, and 41). Even though the integron platform lacks the typical characteristics of bacteriophages (tRNA insertion preference, mobilization genes, terminal flanking direct repeats, and packaging functions), the similarity of the integron system to bacteriophages at the level of the integrase gene, the attachment sites, and the large number of ORFs of unassigned function is intriguing. It is tempting to speculate that the ancestral integron was an immobilized prophage. If so, the integron and phage systems could have diverged to the extent that their functions no longer share any overlap, except for the recombination machinery itself. Although their origin remains elusive, the presence of SIs among diverse Gram-negative bacteria establishes both an ancestry for MRIs and the quintessential role of integrons in adaptive bacterial evolution.

We especially thank Prof. Maurice Hofnung (Unité de Programmation Moléculaire et Toxicologie Génétique) for his unconditional and enthusiastic support. We also thank Stéphane Blanchard for his sequencing skills, Navith Kaim, Latefa Biskri, and Alexandre Prieur for their technical assistance, Dr. Douglas Bartlett for *Photobacterium* and *Vibrio* strains, Drs. Elizabeth Raleigh and George Feller for their gifts of *X. campestris* pv. *badrii* and pULG323, respectively. Thank you to The Institute for Genomic Research and the Department of Energy for making their genome sequence data available to all. In particular, we thank Chantal Bizet and Dominique Clermont of the Collection de l'Institut Pasteur and Dr. Martina Ochs for critical reading of the manuscript and insightful discussions. D.R.-M. was supported by a Manlio Cantarini Postdoctoral Fellowship (Institut Pasteur) and is currently a European Molecular Biology Organization Postdoctoral Fellow. This work was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique, and the Program de Recherche Fondamentale en Microbiologie et Maladies Infectieuses et Parasitaires from the Ministère de l'Éducation Nationale, de la Recherche et de la Technologie.

- Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000) *Nature (London)* **405**, 299–304.
- Davies, J. E. (1994) *Science* **264**, 375–382.
- Sundström, L., Radström, P., Swedberg, G. & Sköld, O. (1988) *Mol. Gen. Genet.* **213**, 191–201.
- Stokes, H. W. & Hall, R. M. (1989) *Mol. Microbiol.* **3**, 1669–1683.
- Ouellette, M., Bissonnette, L. & Roy, P. H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7378–7382.
- Martinez, E. & de la Cruz, F. (1990) *EMBO J.* **9**, 1275–1281.
- Hall, R. M. & Collis, C. M. (1995) *Mol. Microbiol.* **15**, 593–600.
- Rowe-Magnus, D. A. & Mazel, D. (1999) *Curr. Opin. Microbiol.* **2**, 483–488.
- Hall, R. M. & Stokes, H. W. (1993) *Genetica* **90**, 115–132.
- Sundström, L. (1998) *APMIS Suppl.* **84**, 37–42.
- Stokes, H. W., O'Gorman, D. B., Recchia, G. D., Parsekhian, M. & Hall, R. M. (1997) *Mol. Microbiol.* **26**, 731–745.
- Recchia, G. D. & Hall, R. M. (1995) *Microbiology* **141**, 3015–3027.
- Levesque, C., Brassard, S., Lapointe, J. & Roy, P. H. (1994) *Gene* **142**, 49–54.
- Collis, C. M., Kim, M. J., Stokes, H. W. & Hall, R. M. (1998) *Mol. Microbiol.* **29**, 477–490.
- Mazel, D. & Davies, J. (1999) *Cell. Mol. Life Sci.* **56**, 742–754.
- Hall, R. M. & Collis, C. M. (1998) *Drug Resist. Updates* **1**, 109–119.
- Poiré, L., Le Thomas, I., Naas, T., Karim, A. & Nordmann, P. (2000) *Antimicrob. Agents Chemother.* **44**, 622–632.
- Mazel, D., Dychinco, B., Webb, V. A. & Davies, J. (1998) *Science* **280**, 605–608.
- Trucksis, M., Michalski, J., Deng, Y. K. & Kaper, J. B. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 14464–14469.
- Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Umayam, L., et al. (2000) *Nature (London)* **406**, 477–483.
- Rowe-Magnus, D. A., Guerout, A.-M. & Mazel, D. (1999) *Res. Microbiol.* **150**, 641–651.
- Barker, A., Clark, C. A. & Manning, P. A. (1994) *J. Bacteriol.* **176**, 5450–5458.
- Thomson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **24**, 4876–4882.
- Felsenstein, J. (1995) PHYLIP (Dept. of Genetics, Univ. of Washington, Seattle).
- Page, R. D. M. (1996) TREEVIEW (Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, Univ. of Glasgow, Glasgow, U.K.).
- Sirko, A., Zatyka, M., Sadowy, E. & Hulanicka, D. (1995) *J. Bacteriol.* **177**, 4134–4136.
- Feller, G., Thiry, M., Arpigny, J.-L., Mergeay, M. & Gerday, C. (1990) *FEMS Microbiol. Lett.* **66**, 239–244.
- Chambers, S. P., Prior, S. E., Barstow, D. A. & Minton, N. P. (1988) *Gene* **68**, 139–149.
- Mazel, D., Dychinco, B., Webb, V. & Davies, J. (2000) *Antimicrob. Agents Chemother.* **44**, 1568–1574.
- Gravel, A., Messier, N. & Roy, P. H. (1998) *J. Bacteriol.* **180**, 5437–5442.
- Jain, R., Rivera, M. C. & Lake, J. A. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 3801–3806.
- Ochman, H. & Wilson, A. C. (1987) *J. Mol. Evol.* **26**, 74–86.
- Van Cott, E. M. & Wilson, G. G. (New England Biolabs) (1991) U.S. Patent 4983542.
- Martin, C., Timm, J., Rauzier, J., Gomez-Lus, R., Davies, J. & Gicquel, B. (1990) *Nature (London)* **345**, 739–743.
- Jones, M. E., Peters, E., Weersink, A. M., Fluit, A. & Verhoef, J. (1997) *Lancet* **349**, 1742–1743.
- Recchia, G. D. & Hall, R. M. (1997) *Trends Microbiol.* **5**, 389–394.
- Demain, A. L. (1983) *Biochem. Soc. Symp.* **48**, 117–132.
- Ogawa, A. & Takeda, T. (1993) *Microbiol. Immunol.* **37**, 607–616.
- van Dongen, W. M. A. M., van Vlerken, M. M. A. & DeGraaf, F. K. (1987) *Mol. Gen. (Life Sci. Adv.)* **6**, 85–91.
- Cheetham, B. F. & Katz, M. E. (1995) *Mol. Microbiol.* **18**, 201–208.
- Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E. & Hatful, G. F. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2192–2197.